

基于MaxEnt的华山松大小蠹在中国潜在分布区预测

王茹琳 余华彬 王闫利 姜淦 林姗 沈沾红
(四川省农村经济综合信息中心, 成都 610072)

摘要: 华山松大小蠹是华山松的毁灭性害虫, 明确华山松大小蠹在中国的潜在分布对有效控制其扩散蔓延有重要的指导意义。基于最大熵算法的生态位模型MaxEnt和地理信息系统软件DIVA-GIS对华山松大小蠹进行适生区分析及预测, 用受试者工作特征曲线对预测模型和结果进行评估, 用Jackknife法分析影响华山松大小蠹分布的重要因子。结果表明, 华山松大小蠹在我国适生范围比较集中, 高风险区主要分布在陕西南部、四川东北部和甘肃东南部三省交界地区。确定了影响华山松大小蠹分布的主导气候因子, 即最干季度平均温度、最暖季度降水量、降水量变异系数、温度季节性变化标准差和最冷季度降水量。分析了主导气候因子与该虫分布概率的关系, 并对下一步的工作思路进行了介绍。

关键词: 华山松大小蠹, MaxEnt, 适生性分析, 影响因子

DOI: 10.3969/j.issn.2095-1973.2016.05.005

Potential Geographical Distribution of *Dendroctonus armandi* in China Based on the MaxEnt Prediction

Wang Rulin, Yu Huabin, Wang Yanli, Jiang Gan, Lin Shan, Shen Zhanhong
(Sichuan Provincial Rural Economic Information Centre, Chengdu 610072)

Abstract: *Dendroctonus armandi* is a destructive insect of *Pinus armandi*, it caused serious economic losses annually to the forest industry in China. Determination of the potential geographical distribution of *Dendroctonus armandi* is an important factor for effectively controlling the spread of this pest. In this study, the niche model MaxEnt and Diva-gis were employed to analyze and predict the suitable distribution area of *Dendroctonus armandi*, the ROC was used to evaluate the prediction model and the prediction results, and the Jackknife analysis was adopted to analyze the most important environmental factors that affect the occurrence of *Dendroctonus armandi*. The results show that *Dendroctonus armandi* had a concentrated distribution regions: the south of the Shaanxi Province, the northeast of the Sichuan Province and the southeastern Gansu Province. Four dominant climatic factors controlling *Dendroctonus armandi* distribution were determined, including the mean temperature in the driest quarter, the precipitation in the warmest quarter, the precipitation seasonality, the temperature seasonality and the precipitation in the coldest quarter. Here, we analyze the relationship between dominant climatic factors and the probability of the pest presences, and suggest a thinking way for the further research.

Keywords: *Dendroctonus armandi*, MaxEnt, suitability analysis, environmental factors

0 引言

华山松大小蠹 (*Dendroctonus armandi* Tsai and Li) 又名凝脂小蠹, 属鞘翅目 (Coleoptera) 齿小蠹科 (Ipidae) 昆虫, 目前仅在我国有分布。由于该虫隐蔽性强, 世代重叠严重, 防治十分困难, 每年大量发生危害, 严重影响木材材质, 降低经济价值, 甚至导

致华山松大量死亡, 给华山松林区生态环境建设和林业产业带来巨大的损失^[1-2]。近年来, 华山松大小蠹在四川和陕西发生日益严重, 发生面积不断扩大, 为防止该虫进一步扩散和危害, 科学研究并明确其在我国的潜在适生区分布, 制定长治有效的防治措施已经刻不容缓。

Phillips等^[3]研发了一种生态位模型软件MaxEnt (maximum entropy modeling), 它是以最大熵理论 (maximum entropy method) 为基础的一种选择性方法, 用于密度估计和物种分布预测, 采用物种出现点数据和环境变量数据对物种生境适宜性进行评价, 选择熵最大的分布最为最优分布, 推测该物种在目标地

收稿日期: 2014年1月24日; 修回日期: 2014年7月10日
第一作者: 王茹琳 (1986—), Email: wrl_1986_1@163.com
资助信息: 中国气象局“2012年业务专项经费”; 四川省气象局科学技术研究开发课题 (川气课题 2013-开发-10)

区的适生分布, 并分析环境变量对分布概率的影响。物种分布和森林生态类型的主要影响因子往往为气象要素, 气候变化对物种的生理、分布区域及与天敌关系等产生了较大的影响^[4]。

本研究采用MaxEnt, 分析并预测了华山松大小蠹在我国的潜在生物分布区域, 用Jackknife法分析了影响华山松大小蠹分布的重要因子, 为开展该虫在我国的分布区域测报提供技术支持, 为制定合理、有效的检疫及防治措施提供理论依据。

1 材料与方法

1.1 华山松大小蠹分布数据的来源与处理

数据来源: 物种分布数据的获得一般通过实地考察、国内外发表论文和标本馆记录三种方式获得^[5-6]。本研究中, 华山松大小蠹的分布数据主要通过实地考察和查阅国内外公开发表的相关论文获取。

数据处理: 通过Google地图查找分布点的经纬度, 按照MaxEnt的要求, 将华山松大小蠹实际分布点按物种名、分布点经度和分布点纬度顺序储存为csv格式的文件^[7]。

1.2 试验软件

研究使用的生态位分析软件为MaxEnt, 由Phillips等^[3]编写, 工作原理可表述为: 一个物种在没有任何条件约束的情景下, 会尽最大可能扩散蔓延, 接近均匀分布^[8]。模型计算基于贝叶斯定理, 利用Gibbs分布将特征集进行加权并作为参数, 进行一系列运算, 得到物种分布的最大熵联合分布估计, 进而建立地理分布预测模型^[9]。软件获取方法: 提供研究者姓名、研究机构信息和电子邮件地址后, 可以从<http://www.cs.princeton.edu/~schapire/maxent/>免费下载, 本研究使用的版本为3.3.3。

用于地图处理的GIS软件采用免费软件DIVA-GIS, 从<http://www.diva-gis.org/>下载, 版本为7.5。

1.3 环境层数据和地图数据

环境层数据: 采用WORLDCLIM (<http://www.worldclim.org/>) 免费下载的1950—2000年19个生物气候变量的监测数据, 数据空间分辨率为5'。19个气候变量为: 年平均气温 (bio1)、平均日较差 (bio2)、等温性 (bio3)、温度季节性变化标准差 (bio4)、最暖月最高温度 (bio5)、最冷月最低温度 (bio6)、年均温变化范围 (bio7)、最湿季度平均温度 (bio8)、最干季度平均温度 (bio9)、最暖季度平均温度 (bio10)、最冷季度平均温度 (bio11)、年降水量 (bio12)、最湿月降水量

(bio13)、最干月降水量 (bio14)、降水量变异系数 (bio15)、最湿季度降水量 (bio16)、最干季度降水量 (bio17)、最暖季度降水量 (bio18) 和最冷季度降水量 (bio19)。

地图数据: DIVA-GIS所需的分析底图采用从国家基础地理信息系统 (<http://nfgis.nsd.gov.cn/>) 下载的1:400万的中国行政区划图。

1.4 研究方法

MaxEnt模型操作方法^[10-11]: 将华山松大小蠹的分布数据和19个生物气候变量导入MaxEnt, 参数设置为: 25%的分布点作为测试集, 剩余的分布点为训练集, 选择刀切法分析影响因子, 输出格式为ASCII。图形处理: ASCII格式文件导入DIVA-GIS进行格式转换, 将华山松大小蠹在我国的潜在适生区划分为4个等级: 高风险区 (61~100, 红色)、中风险区 (31~60, 橙色)、低风险区 (11~30, 黄色) 和无风险区 (0~10, 白色)。

主导因子的选取: 根据19个生物气候因子对华山松大小蠹分布影响贡献率的大小, 选出影响华山松大小蠹分布的主导因子。在此基础上, 重建华山松大小蠹在中国分布的最大熵模型, 并进行模拟结果精度评价^[12]。

2 结果与分析

2.1 华山松大小蠹在中国的适生性分析

利用DIVA-GIS软件和MaxEnt生态位模型分析处理的华山松大小蠹在中国的适生性分布结果如图1所示: 华山松大小蠹在我国的高风险区主要集中在陕西

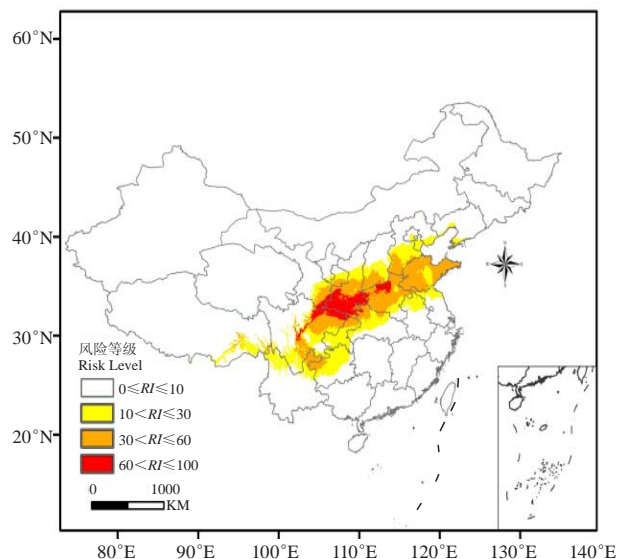


图1 基于MaxEnt的华山松大小蠹在我国的适生区域
Fig. 1 Map of potential regions suitable for *Dendroctonus armandi* in China, based on the MaxEnt prediction

南部、四川东北部和甘肃东南部三省交界地区，以及湖北西北部地区。根据实际调查及查阅文献得知，该区域也是该虫近年来发生严重的地区。中风险区分布范围以高风险区为核心向外扩大，主要集中在四川、陕西、山西、河南和山东等省。

2.2 华山松大小蠹发生的影响因子分析及 ROC 曲线评价

图2为影响华山松大小蠹发生的主要气候变量的刀切分析图，从图中可以看出：最干季度平均温度（bio9）对华山松大小蠹分布影响最大，训练增益（training gain）超过了2.0。最冷季度平均温度（bio11）、最湿月降水量（bio13）和最冷月最

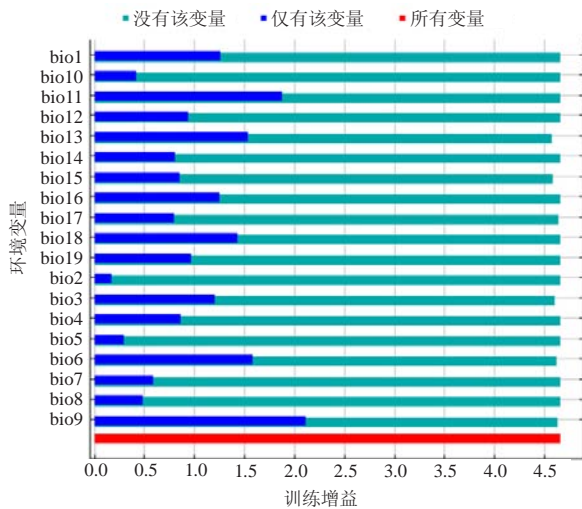


图2 影响华山松大小蠹发生的环境变量的刀切分析
Fig.2 The jackknife analysis of importance of environmental variables for *Dendroctonus armandi*

2.3 主导气候因子筛选及其与存在概率关系

表1给出了影响华山松大小蠹在中国分布的前5个主要气候因子的贡献百分率和累计贡献百分率。按照贡献百分率由大到小排序依次为：最干季度平均温度（25.4%）、最暖季度降水量（21.3%）、降水量变异系数（11%）、温度季节性变化标准差（10.1%）和最

表1 影响华山松大小蠹分布的气候因子贡献百分率和累积贡献百分率

Table 1 The accumulated contribution of each environmental variable to the potential distribution of *Dendroctonus armandi* resulted from the MaxEnt

环境变量	MaxEnt 贡献率 (%)	累计贡献率 (%)
最干季度平均温度	25.4	25.4
最暖季度降水量	21.3	46.7
降水量变异系数	11	57.7
温度季节性变化标准差	10.1	67.8
最冷季度降水量	9.9	78.7

低温度（bio6）对该虫分布影响较大，训练增益均超过1.5。平均日较差（bio2）、最暖月最高温度（bio5）、最湿季度平均温度（bio8）和最暖季度平均温度（bio10）对华山松大小蠹分布影响较小，它们的训练增益都小于0.5。

目前研究中，一般采用ROC曲线分析方法对物种潜在分布预测模型进行评价^[13-14]，它通过将连续变量设定出多个不同的临界值，从而计算出一系列敏感性和特异性，AUC值在0~1，值越大表示模型预测效果越好。本研究所得的AUC值为0.995（图3），比随机分布模型的AUC值（约为0.5）大，说明本实验的预测结果具有较高的可信度。

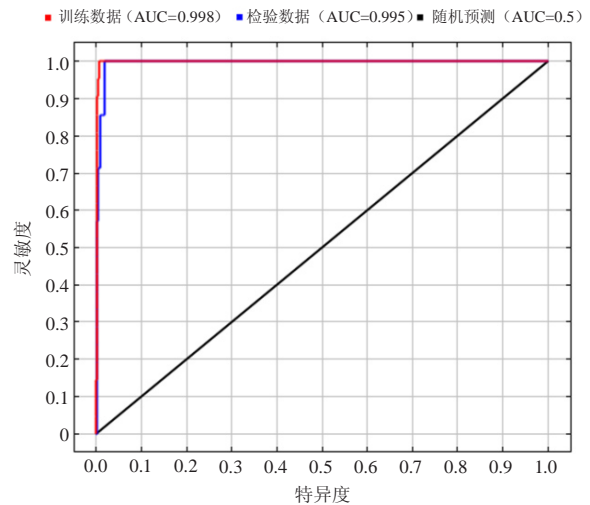


图3 ROC曲线及AUC值
Fig. 3 The ROC curve and AUC value

冷季度降水量（9.9%）。这5个气候因子的累积贡献率达到78.7%，可以认为这5个气候因子是影响华山松大小蠹分布的关键主导因子。

图4反映了5个主要气候因素对华山松大小蠹分布的影响。图4a曲线显示出当最干季度平均温度约0~4℃时，华山松大小蠹分布概率均超过0.6，达到2℃时的分布概率为最大值。图4b曲线显示出最暖季度降水量在约200~1000mm时，华山松大小蠹分布概率变化显著，在约300mm时上升至最大值0.7，降水量超过500mm后分布概率迅速下降，分布概率小于0.5。图4c曲线显示出分布概率与降水量变异系数呈负相关，小于约30时，分布概率超过0.9，随着变异系数的增加，分布概率逐渐降低。图4d曲线显示出温度季节性变化标准差值在约7500时，分布概率达到最大值。图4e曲线显示出最冷季度降水量在300~500mm时，

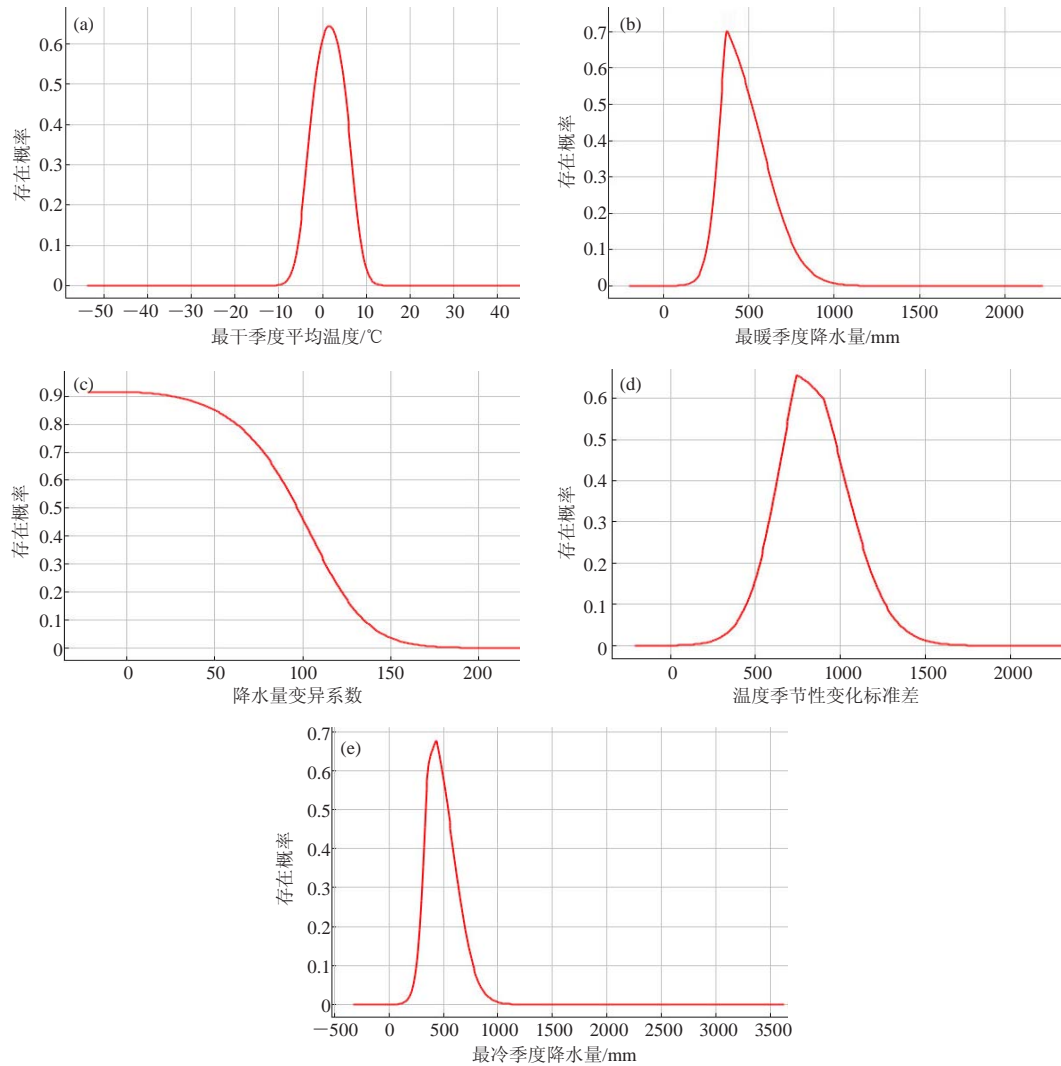


图4 华山松大小蠹分布概率对5个重要气候因子的响应曲线

Fig. 4 Response curves of the *Dendroctonus armandi* to environmental variables resulted from the MaxEnt models

分布概率超过0.6。

3 讨论

MaxEnt模型是一种基于机器学习和数学统计的生态位模型，是目前常用的物种适生区预测软件，相对于其他一些常用预测模型有其独特的优势^[15]。本研究使用DIVA-GIS软件将MaxEnt的预测结果进行分级显示，直观地表示出华山松大小蠹在我国的潜在适生区。

研究表明：1) 华山松大小蠹发生的高度适生区，即高风险区，主要集中在我国的四川、陕西、甘肃和湖北等地，而这些区域也是该虫时常爆发成灾的地区，说明预测结果有较高的可信度；2) MaxEnt模型的AUC值为0.995，表明预测效果非常好；3) 分析各生物气候因子对华山松大小蠹潜在地理分布的贡献率，筛选出了影响华山松大小蠹分布的主导气候因子，即最干季

度平均温度、最暖季度降水量、降水量变异系数、温度季节性变化标准差和最冷季度降水量。其中，最干季度平均温度训练增加超过2.0，表明此因子对华山松大小蠹分布影响最大，此结果有助于深入理解华山松大小蠹与气候变化的关系，对于科学的制定华山松大小蠹的防控措施具有重要的意义。

四川省农村经济综合信息中心2011年开展了基于本体知识库的气象灾害实时预警分析与生成，应用于气象服务客观自动化分析。在中国天气网四川站和成都公共气象服务网开通了耕作气象、山地气象、灾害气象等栏目，取得了良好的运行效果^[16-17]。本研究分析了影响华山松大小蠹分布的主导气候因子，并得到了各影响因子与存在概率的分布曲线，计划在下一步工作中，借鉴气象灾害本体设计的思路，针对主导气候因子进行风险等级的划分，将得出的风险等级数据

经过地理信息系统处理，可以得到有关华山松大小蠹的近实时入侵风险本体分析图。目前这部分研究还处于探索阶段，还有大量工作有待完成。

通过本研究，对华山松大小蠹在全国的潜在适生分区有了明确直观的了解，为避免该害虫的进一步扩大，林业部门在华山松大小蠹高度适生区域（四川、陕西和甘肃等省份）应进一步加强预测预报，建立针对该虫的更加完善的监测预警体系，注意加强检疫，通过科学有效的预测及防治手段，把该虫的发生风险控制发生在发生阈值以下，这样才能更加有效的保护我国的华山松林。

参考文献

- [1] 王茹琳, 杨伟, 杨佐忠, 等. 华山松大小蠹对9种植物挥发物的EAG和行为反应. 中国森林病虫, 2011, 30(1): 23-26.
- [2] 王茹琳, 杨伟, 杨佐忠, 等. 华山松大小蠹对几种寄主挥发物组成的EAG和行为反应. 生态学杂志, 2011, 30(4): 724-729.
- [3] Phillips S J, Dudik M, Schapire R E. A maximum entropy approach to species distribution modeling. Proceedings of the 21st International Conference on Machine Learning. New York: ACM Press, 2004.
- [4] 张建新, 包云轩, 李芬, 等. 林业有害生物发生发展与气象条件关系研究进展. 世界林业研究, 2010, 23(1): 33-38.
- [5] 曲伟伟, 李志红, 黄贵修, 等. 利用MAXENT预测橡胶树棒孢霉落叶病在中国的适生区. 植物保护, 2011, 37(4): 52-57.
- [6] 曾辉, 黄冠胜, 林伟, 等. 利用MaxEnt预测橡胶南美叶疫病菌在全球的潜在地理分布. 植物保护, 2008, 34(3): 88-92.
- [7] 房锋, 张朝贤, 黄红娟, 等. 基于Maxent的麦田恶性杂草节节麦的潜在分布区预测. 草业学报, 2013, 22(2): 62-70.
- [8] Jaynes E T. Information theory and statistical mechanics. Physical Review, 1957, 106(4): 620-630.
- [9] 殷晓洁, 周广胜, 隋兴华, 等. 辽东栎林潜在地理分布及其主导. 林业科学, 2013, 48(8): 10-14.
- [10] Brotons L, Thuiller W, Araujo M B, et al. Presence-absence versus presence-only modeling methods for predicting bird habitat suitability. Ecography, 2004, 27(4): 437-448.
- [11] Elith J, Graham C H, Anderson R P, et al. Novel methods improve prediction of species distributions from occurrence data. Ecography, 2006, 29(2): 129-151.
- [12] 殷晓洁, 周广胜, 隋兴华, 等. 蒙古栎地理分布的主导气候因子及其阈值. 生态学报, 2013, 33(1): 103-109.
- [13] Berger A L, Della Pietra S A, Della Pietra V J. A maximum entropy approach to natural language processing. Computational Linguistics, 1996, 22(1): 39-71.
- [14] 邢丁亮, 郝占庆. 最大熵原理及其在生态学中的应用. 生物多样性, 2011, 19(3): 295-302.
- [15] 陈晨, 龚伟荣, 胡白石, 等. 基于地理信息系统的红火蚁在中国适生区的预测. 应用生态学报, 2006, 17(11): 2093-2097.
- [16] 何险峰, 雷升锴, 张祥锋, 等. 本体化气象服务实现与应用. 计算机应用, 2012, 32(S2): 226-231.
- [17] 何险峰, 张祥锋, 郑利娟, 等. 气象灾害本体设计. 气象科技, 2012, 40(6): 1007-1012.

《应用气象学报》创刊30周年纪念刊出版

本刊编辑部

2016年9月，由中国气象科学研究院等多家单位联合主办的大气科学综合性学术期刊《应用气象学报》出版了创刊30周年纪念刊。杂志刊载了11篇综述性论文，内容涵盖了降水预报、农业气象、高性能计算机与气象卫星的应用、污染观测等多个领域的研究进展。

纪念刊收录的文章和主要作者如下。

中国冬季积雪特征及欧亚大陆积雪对中国气候影响——张人禾等

适应气候变化的国际行动和农业措施研究进展——周广胜等

定量降水预报技术进展——毕宝贵等

高性能计算技术在气象领域的应用——赵立成等

华南暖区暴雨研究进展——何立富等

雷暴闪电放电活动对电离层影响的研究进展——张义军等

中国南方旱涝年际年代际变化及成因研究进展——李维京等

气象卫星可见光红外光学成像仪发展沿革——杨忠东等

我国霾和光化学污染观测研究进展——徐晓斌

农业气象灾害监测预测技术研究进展——郭建平

江淮流域持续性极端降水及预报方法研究进展——翟盘茂等